

Making Sense of Transcription Networks

Trevor R. Sorrells¹ and Alexander D. Johnson^{1,*}

¹Department of Biochemistry & Biophysics and Department of Microbiology & Immunology, Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

*Correspondence: ajohnson@cgl.ucsf.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.014>

When transcription regulatory networks are compared among distantly related eukaryotes, a number of striking similarities are observed: a larger-than-expected number of genes, extensive overlapping connections, and an apparently high degree of functional redundancy. It is often assumed that the complexity of these networks represents optimized solutions, precisely sculpted by natural selection; their common features are often asserted to be adaptive. Here, we discuss support for an alternative hypothesis: the common structural features of transcription networks arise from evolutionary trajectories of “least resistance”—that is, the relative ease with which certain types of network structures are formed during their evolution.

Introduction

The complexity of cells continues to fascinate scientists. Two broad views are often advanced to account for such complexity. In one, it is assumed that any complexity must necessarily benefit the cell. Some cell and molecular biologists go even further and discuss how a particular mechanism was “designed” by evolution to be perfectly matched to its task. As with a machine, it is assumed that every molecular nut and bolt must have a purpose. Because this view seems intuitive and relatively simple (after all, examples abound of animals, plants, and microbes adapted to their environments), it is often invoked to explain any aspect of cell and molecular biology. A different view, the one we elaborate here, is embodied in Dobzhansky’s famous line, now a cliché, “nothing in biology makes sense except in the light of evolution.” According to this view, any rationalization of a modern cellular mechanism depends critically on understanding its evolutionary history. We argue that this emphasis on evolutionary history is especially appropriate for analyzing transcription circuits and for rationalizing their structures. This view has explanatory power in that it can readily account for some of the more bewildering and counterintuitive features of modern transcription circuits; it also gives us insight into the best ways to describe and study such circuits.

In this Perspective, we first review common features of transcription network structures—observed across diverse species—and argue that these similarities cannot be the result of descent from a single ancestral circuit possessing these characteristics. Next, we consider key biochemical and biophysical properties of transcription regulators and *cis*-regulatory sequences that make certain evolutionary pathways much more probable than others, in part because they circumvent fitness barriers. Finally, we argue that many aspects of transcription circuits, particularly those that seem overly complex and counterintuitive, can be understood as relatively crude products of high-probability evolutionary trajectories rather than as highly optimized, specific solutions.

The arguments discussed in this perspective rely heavily on prior ideas advanced by evolutionary biologists, particularly

those ideas concerning the role of non-adaptive mutations in generating complexity (Covello and Gray, 1993; Doolittle, 2013; Force et al., 1999; Gray et al., 2010; Lukeš et al., 2011; Lynch, 2007a, 2007b, 2014; Stoltzfus, 1999; Zuckerkandl, 1997). Although sometimes dismissed as unimportant (or uninteresting), non-adaptive mutations can have a profound role in generating evolutionary novelty. Of particular importance is the idea, sometimes called “constructive neutral evolution,” that changes that arise neutrally can open up new evolutionary pathways; in some cases, changes that arose non-adaptively can become essential for function if they are incorporated into subsequent layers of evolutionary change. Through this sequence of events, molecular and organismal complexity can be increased through non-adaptive mutations. As we discuss, the biochemical and biophysical properties of transcription network components support the idea that their evolutionary trajectories—which depend on mutation, selection, and genetic drift—lead to specific types of structures. Because their components are highly conserved across eukaryotes, we argue that it is inevitable that networks across a wide variety of species tend to converge on similar structures. We propose that these common structures are not likely to represent optimized solutions but are, in a sense, “default” evolutionary products.

Depictions of Transcription Networks

For the most part, genome-wide studies of transcriptional network structures have been largely descriptive, often culminating in large “hairball” diagrams such as those depicted in Figure 1. Their complexity has made it difficult to formulate simple conclusions regarding the logic or outputs of these networks, particularly since quantitative parameters and dynamic measurements are typically lacking.

Although there are many components of gene expression networks, we will focus here on only two key elements, transcription regulators and *cis*-regulatory sequences. We define transcription regulators as sequence-specific DNA-binding proteins that control the transcription of specific genes by binding to *cis*-regulatory sequences, short (typically 6–15 nucleotides) DNA

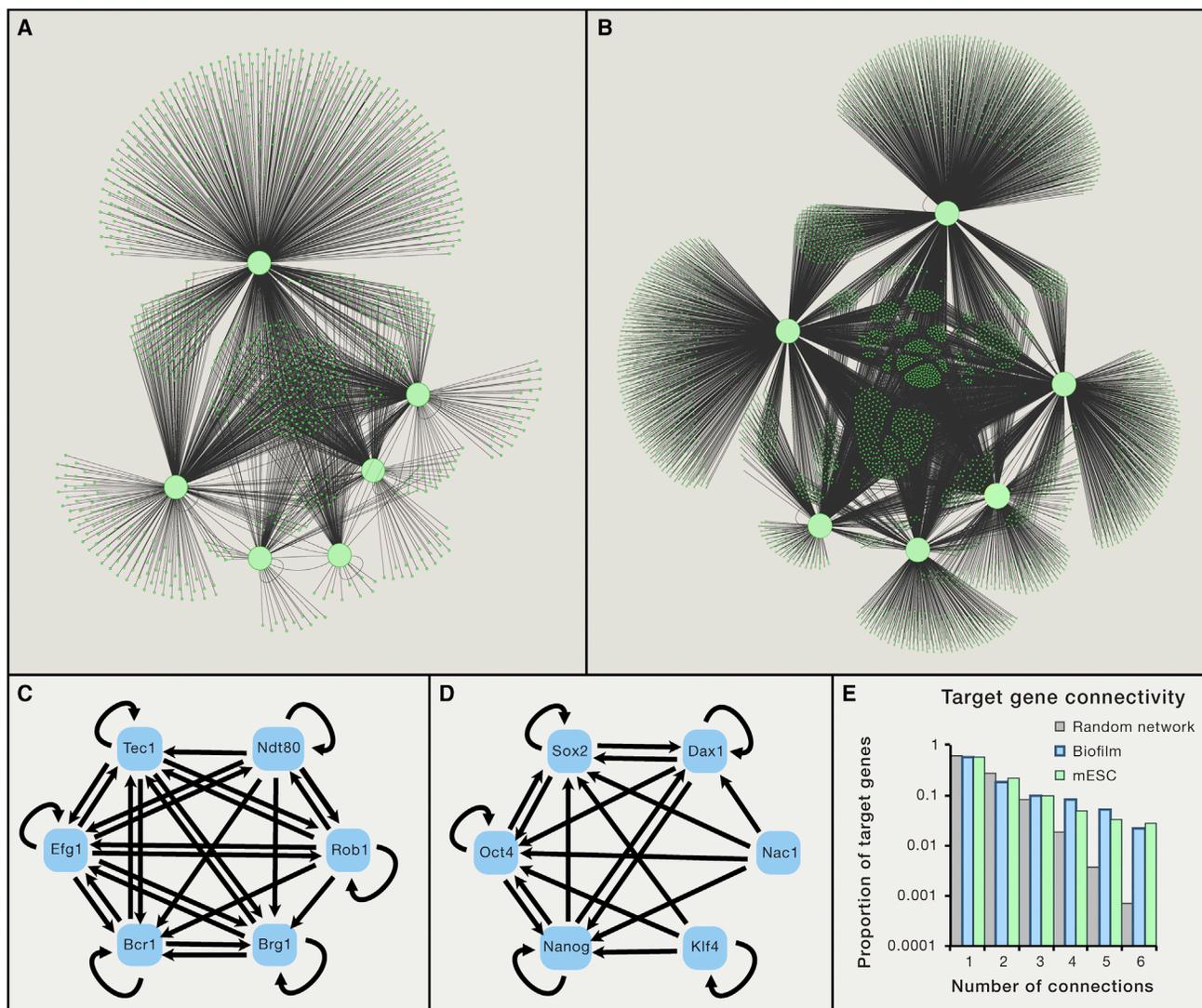


Figure 1. Typical Depictions of Transcription Regulatory Networks

(A and B) (A) The *C. albicans* biofilm network (Nobile et al., 2012) and (B) the *M. musculus* embryonic stem cell network (Kim et al., 2008) are depicted as graphs where balls represent genes and lines represent the binding of transcription regulators to intergenic regions. Master transcription regulators (defined in the text) are shown as large balls, and “target genes” are shown as small balls. For the stem cell network, only the six most heavily connected transcription regulators are shown.

(C and D) Close-up of the core of each network, showing only the binding connections between the master transcription regulators. Directionality of the connection is indicated by arrows. Note that the arrows refer only to binding connections and do not imply that the connection activates the recipient gene. (C) *C. albicans* biofilm, (D) mouse stem cell networks.

(E) The degree of connectivity for nodes in the two networks. The two biological networks show a larger proportion of nodes with high connectivity than would be found in a random network (Lee et al., 2002).

sequences. It is the distribution of these *cis*-regulatory sequences across the genome that largely specifies the time, place, and rate of each gene’s transcription; this information is “read” by transcription regulators, whose binding to DNA specifies, often through a complex series of downstream steps, the rate of transcription of the gene. Although in many eukaryotic species, *cis*-regulatory sequences are typically located within several thousand nucleotide pairs of the genes they control, in plants and animals, they can be spread out over hundreds of thousands of nucleotide pairs. Nearly all eukaryotic genes are directly controlled by more than one transcription regulator,

and most genes respond to dozens of regulators, specified by the identity and arrangement of their *cis*-regulatory sequences. We also know, from decades of “promoter bashing” experiments, that *cis*-regulatory sequences can be moved from one gene to another (and from one species to another) and still retain much of their specificity to direct transcription. Finally, transcription regulators typically bind cooperatively to DNA, a fundamental property that, as we shall discuss, has important implications for network evolution.

Many additional proteins besides transcription regulators are needed to transcribe a gene (for example, RNA polymerase

Table 1. Metrics Comparing *C. albicans* Biofilm and Mouse Embryonic Stem Cell Networks

	Biofilm	mESC
Master transcription regulators	6	6
Connections	2018	7234
Target genes	1037	3968
Fraction of genome in network	0.17	0.21
Binding feed-forward loops	3145	6886
Non-functional binding events	1207	unknown

Connections were determined by whole-genome chromatin immunoprecipitation and the table values are simply total counts without regard for the number of significant figures. As such, they should be regarded as approximate, particularly since there is no evidence that either circuit is completely described. “Non-functional binding” was defined as genes whose expression does not change when a direct regulator is deleted from the genome.

and chromatin remodeling complexes), but the specification of which genes are transcribed in which tissues is determined by direct binding connections between transcription regulators and genes (or more precisely, the *cis*-regulatory sequences of that gene). This information is summarized in diagrams such as those in Figure 1.

If a given transcription regulator occupies the *cis*-regulatory sequences associated with a gene in vivo (as determined, for example, by a chromatin immunoprecipitation experiment), we will refer to that gene as a target gene of the transcription regulator. We realize that this convention does not require that the binding of the regulator to DNA be proven to be functional in the organism. There are three reasons for nonetheless including these connections in diagrams such as those in Figure 1. (1) The “function” of a given connection has been demonstrated in only a small number of cases; for the great majority of reliable binding data, no direct test has been performed. (2) Although many approaches (e.g., conservation across species or experimental mutation of the *cis*-regulatory sequence) can provide strong evidence for function, it is impossible to rigorously establish that a binding connection is non-functional under all possible conditions. (3) The DNA-binding properties of transcription regulators predict that, in vivo, there will be some degree of non-functional binding (Lin and Riggs, 1975). Such “non-functional” binding events are nonetheless real properties of evolving transcription networks.

Depictions of transcription networks based on these conventions often show “master transcription regulators” and target genes as nodes (balls) and regulatory interactions as edges (lines) between these nodes (Figures 1A and 1B). Although the term master transcription regulator is used in many different ways in the literature (Chan and Kyba, 2013), we define it, for the purpose of this article, as a transcription regulator (1) whose presence is required to carry out the specific biological process controlled by the network and (2) whose ectopic expression alone or in combination with other regulators can trigger the biological process even in the absence of the ordinary developmental or environmental signals (Halder et al., 1995; Takahashi and Yamanaka, 2006; Tapscott et al., 1988; Tursun et al., 2011; Vierbuchen et al., 2010; Zordan et al., 2007).

Common Features of Transcription Networks

We first compare two transcription networks from two different species that coordinate two different biological processes, but were deduced by similar methodologies. The network specifying the embryonic stem cell state (pluripotency) was chosen because it has been studied extensively by numerous labs and is supported by multiple studies (Boyer et al., 2005; Kim et al., 2008). For comparison, we chose the circuit controlling biofilm development in the pathogenic yeast *C. albicans*, a network this lab has studied extensively (Nobile et al., 2012). The two networks are depicted in Figures 1A and 1B using a similar graphical format.

These two circuits were chosen, in part, because they might be expected from first principles to have little in common. Mammals and yeast diverged from a common ancestor ~1.5 billion years ago (Wang et al., 1999), and there is little conceptual similarity between biofilm formation and pluripotency. Moreover, the two networks appear to have evolved independently, well after the two lineages split (see below). Yet, the overall structures of the two networks, as depicted in the figure, appear remarkably similar. Both *C. albicans* biofilm development and mouse embryonic stem cell pluripotency are controlled by a set of master transcription regulators that form binding connections among themselves (Figures 1C and 1D) and to the regulatory regions of over a thousand target genes, with multiple master regulators typically binding to the same targets (Figures 1A, 1B, and 1E and Table 1). In both cases, a substantial proportion of the target genes are other transcription regulators, indicating substantial indirect regulation of additional genes. The *C. albicans* genome is significantly smaller than the mouse genome, yet each network comprises about one-fifth of the coding genes in their respective genomes.

Although the two networks control very different processes, their master regulators have similar properties. In both networks, these regulators contain ordinary sequence-specific DNA-binding domains such as homeodomains, MADS domains, and zinc fingers (Weirauch and Hughes, 2011). In some cases, the *cis*-regulatory sequence recognized by a given transcription regulator has not changed significantly since the divergence of yeast and mammals (Hayes et al., 1988). Moreover, transcription regulators from one species (e.g., Gal4 from brewer’s yeast) can control transcription in many different species (e.g., Fischer et al., 1988; Kakidani and Ptashne, 1988). A master regulator for a given process is therefore distinguished by its connections, not its structure.

Key aspects of the *C. albicans* biofilm circuit were formed well after *C. albicans* diverged from closely related, non-pathogenic yeasts (Nobile et al., 2012), providing additional support for the conclusion that the structure of the yeast and mouse networks evolved independently—even though the transcription regulators that would become master regulators for these two networks were present in the common ancestor of both species. These ideas are consistent with the generalization that, although the transcription regulators and their recognition sequences are often deeply conserved, transcriptional networks themselves are rewired at a rapid pace during evolution (reviewed in Li and Johnson, 2010; Tuch et al., 2008b; Weirauch and Hughes, 2010; Wray et al., 2003). Like most generalizations in biology, this one has important exceptions. See, for example, Baker et al. (2011) and

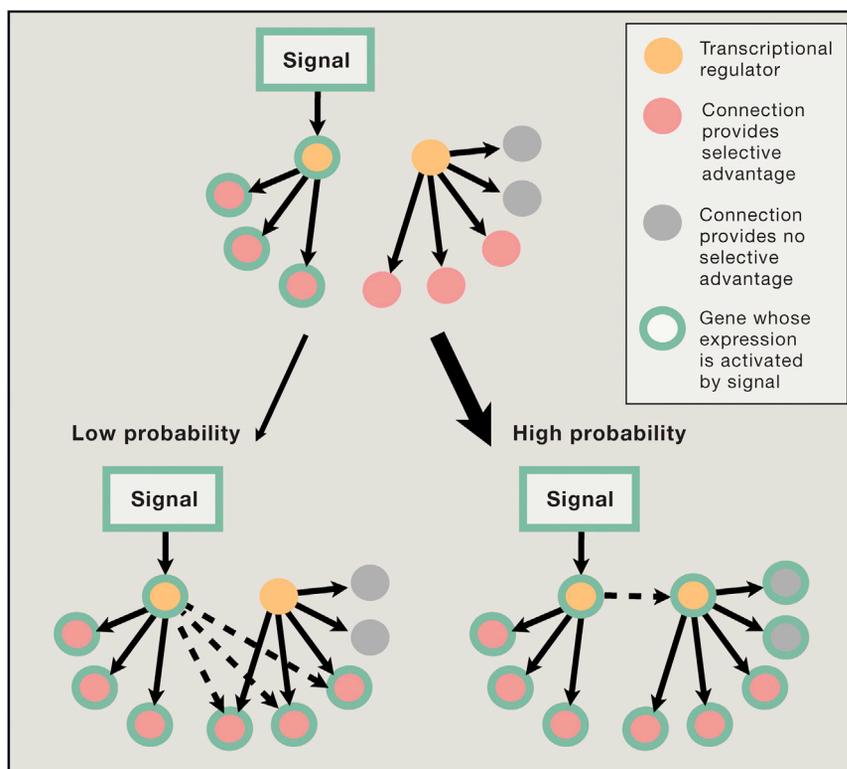


Figure 2. Pathways for Evolving a New Transcriptional Response to a Signal

In this hypothetical scenario, incorporation of three additional genes into the signaling pathway confers a selective advantage. Two alternative paths are possible: (1) the genes could be incorporated one by one through independent changes in their *cis*-regulatory sequences (gene-by-gene model). (2) The new genes could be incorporated through a single incorporation of the transcription regulator that already controls them (regulator-first model). If the incorporation of multiple target genes is needed to confer an increase in fitness, gain of regulation of the transcription regulator will be more probable than the gain of each individual target. As the number of target genes increases, the difference in probability will be greater. Note that the second scenario will likely incorporate additional genes non-adaptively.

virtually all other organisms), and we suggest that the available trajectories for evolutionary change are based on these simple properties coupled with the avoidance of fitness barriers. According to this view, the similarities among independently derived transcription networks arise primarily from the “low-energy” pathways of evolution—rather than the selective pressures specific to one circuit

Sayou et al. (2014) for cases in which the DNA-binding specificity of a regulator has changed dramatically over relatively short periods of evolutionary history. In any case, it is highly unlikely that any of the connections between regulators and target genes in the mouse pluripotency and yeast biofilm networks are conserved from a common ancestor, despite the deep conservation of the DNA-binding properties of the master transcription regulators.

If transcription networks evolve rapidly, why do the embryonic stem cell and biofilm networks appear structurally similar? One hypothesis is that elaborate and interconnected networks such as these represent optimized solutions for organizing biological processes. According to this view, the similarities between these networks result primarily from selection and reflect the same underlying requirements for transcriptional logic—for example, modularity or robustness. Some features of the circuits (for example, the large number of direct and indirect feedback loops) may well reflect these requirements in a general way, but the similarities seem too great to be readily explained this way. We propose instead that circuit architecture is dominated by severe constraints on the evolutionary trajectories available for network evolution. Allowable trajectories, we argue, must (1) be probable from a biochemical and biophysical standpoint and (2) avoid fitness barriers; that is, the allowable trajectories will typically not pass through stages in which the circuit becomes broken and non-functional (Carroll, 2008; Stern and Orgogozo, 2009; Wagner, 2003).

The components of circuits (DNA-binding proteins and *cis*-regulatory sequences) and their properties (for example, cooperative binding) are common to fungi and mammals (as well as

or another. In the following sections, we examine specific properties of networks in more detail and consider the extent to which this idea can account for them.

Size Accrues

One surprising feature of many transcriptional networks is their large size (Borneman et al., 2006; Hernday et al., 2013; Iyer et al., 2001; Junion et al., 2012; Kim et al., 2008; Liang and Biggin, 1998; MacArthur et al., 2009; Mastick et al., 1995; Nobile et al., 2012; Novershtern et al., 2011). As mentioned above, the yeast biofilm network and the mouse embryonic stem cell network, as depicted in Figure 1, incorporate approximately one-fifth of the protein-coding genes in their respective genomes. Although a few examples have been described in which eukaryotic transcription networks appear small (e.g., the mating type specification circuit [Galgoczy et al., 2004] and the galactose regulatory circuit [Ren et al., 2000], both from *S. cerevisiae*), the majority of networks that have been carefully studied using full-genome methods appear larger and more complex than might have been expected.

Why is the typical network so large? In contrast to a model in which every connection in a network serves a specific function in that network, we propose that many target genes in networks are incorporated non-adaptively during the formation of the network. Figure 2 shows a hypothetical example in which a new response to a signal evolves under selection. If there is an advantage of gaining regulation of multiple target genes in response to the signal, it is much more probable to gain a binding site upstream of a single transcription regulator of those genes than to gain binding sites for each individual target gene (Gerhart and

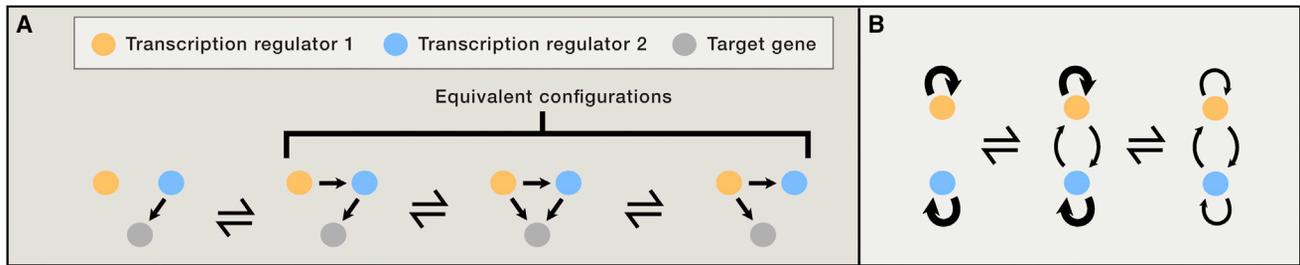


Figure 3. The Tendency for Co-Expressed Regulators to Become Interconnected

(A) Once the orange regulator gains control of the blue regulator, causing them to be expressed at the same time, target genes can, through neutral evolution, rapidly gain and lose binding sites for the two regulators.

(B) Two regulators expressed at the same time each have positive feedback. Subsequently, neutral gains of reciprocal regulation between the regulators can occur while preserving the overall positive feedback control. Over evolutionary time, positive feedback distributed over both regulators (rather than purely autonomous loops) is predicted to occur.

Kirschner, 1997; Raff and Kaufman, 1983). Moreover, because most proteins work in groups, any selective advantage of incorporating a new gene into an existing circuit may not be realized until several genes are brought into the circuit, making the “gene-by-gene” model even less probable. The “regulator-first” model would result in a new regulator being incorporated into the old circuit, along with all the pre-existing target genes of this regulator. Some of these target genes may be extraneous with respect to the new circuit, but, if the original function of the regulator is retained, these connections would nonetheless be maintained by purifying selection. According to this simple idea, newly formed networks would be expected to contain connections nonessential to that network and would therefore be predicted to be larger than strictly necessary.

Experimental evidence suggests that the “regulator-first” scenario is common; that is, networks often form by incorporating new regulators rather than by incorporating individual target genes (Frankel et al., 2012; Monteiro, 2012; Pires et al., 2013). For example, the red wing color in *Heliconius* butterflies takes place through repeated rewiring of the expression pattern of the transcription regulator *optix* rather than one-by-one incorporation of individual target genes (Reed et al., 2011). A second example is found in networks regulating morphological transitions in different yeast species; the regulator Tec1 and its target genes have been incorporated into environmental response networks multiple times (Mösch and Fink, 1997; Nobile et al., 2012; Schweizer et al., 2000). Thus, the regulators-first model of transcription network formation is predicted to lead to the expansion of circuit size beyond that strictly required for the new response. Although this model might be expected to create detrimental pleiotropic effects of expressing many extraneous genes at once, modeling and experimental evidence suggests that this pleiotropy can be alleviated gradually over time (Pavlicev and Wagner, 2012; Qian et al., 2012) or even avoided altogether (Stern and Orgogozo, 2009). It is important to note that these observations, although widely applicable to many eukaryotic species (particularly those with small effective population sizes), do not hold in every case. For example, in some viruses, genomes are compact, regulatory networks are small, and each component and connection contributes to the function of the circuit (Little, 2010).

Gains in Interconnectedness

Another common feature of transcription networks across diverse species is the degree of connectivity between different transcription regulators and between these regulators and their targets (Borneman et al., 2006; Boyle et al., 2014; Junion et al., 2012; Kim et al., 2008; MacArthur et al., 2009; Nobile et al., 2012; Novershtern et al., 2011; Reece-Hoyes et al., 2013). We define this degree as the number of connections made between the master transcription regulators and a given target gene. For example, if a given target gene in the *C. albicans* biofilm network is bound by three different master transcription regulators, the degree of connection of that target gene is three. The degree distributions for the yeast and mouse cases show a similar profile (Figure 1E), one that shows a higher degree of connection than would be predicted for a randomly distributed network (Featherstone and Broadie, 2002; Guelzim et al., 2002).

Rather than speculating what this high degree of interconnectedness might “do for the cell,” we subscribe to the simpler hypothesis that it results from the neutral (i.e., non-adaptive) gains of regulatory connections that inevitably occur over time, particularly in small populations (Lynch, 2007a; Stone and Wray, 2001). This idea can be explained by considering a simple situation, one that would be predicted to arise often by the “regulator-first” model (Figure 3A). Here, one transcription regulator (blue) regulates the target gene (gray). In the regulator-first model, a second transcription regulator (orange) gains control of the blue regulator, and indirectly, the gray target gene. Next, the interconnectedness in this simple scheme would increase if the orange regulator gained direct control of the gray target gene without losing its initial connection.

Why would this happen? It has been argued, using population genetic models, that such connections are predicted to form non-adaptively—that is, without selection for improvement of the circuit (Lynch, 2007a). According to this view, the additional connections do not disrupt the existing regulation, and they arise through random mutations that produce a new DNA-binding site for a transcription regulator. Thus, this increase in total number of connections is predicted to occur, in essence, because nothing stops it. Such a change, even though it arose non-adaptively, can become necessary if subsequent evolutionary changes in the network render its loss detrimental.

Although it might seem counterintuitive that new circuit connections can form non-adaptively, the biochemical features of transcription regulators and *cis*-regulatory sequences predict this. As has been pointed out many times, because *cis*-regulatory sequences are usually short and somewhat degenerate, there is a significant probability that new point mutations will readily create bona fide *cis*-regulatory sequences for existing transcription regulators. Given that target genes often have long intergenic regions in which *cis*-regulatory sequences can function, many target genes would be predicted to develop multiple connections (Lynch, 2007a; Paixão and Azevedo, 2010; Stone and Wray, 2001). Although these additional binding sites may not be under purifying selection (unless the original connection is lost or some other change in the network renders their loss detrimental), they would be predicted to form at a high enough frequency to ensure an appreciable steady-state level of such connections, despite the losses due to mutation.

These same forces are also predicted to lead to the high interconnectedness observed between the master regulators themselves (Figure 1B). Many transcription regulators control their own transcription (Bateman, 1998; Kielbasa and Vingron, 2008; Lee et al., 2002). When two such regulators function at the same time and place (although not necessarily in the same biological function), over time they may acquire regulation of each other through the gain of *cis*-regulatory sequences (Figure 3B). This reciprocal regulation would be redundant (at least in a general sense) with the auto-regulation of each of the transcription regulators themselves and could partially replace it over time, resulting in interlocking, auto-regulatory master regulators of the type we see in Figure 1.

Various types of simulations both support these ideas and highlight additional features that promote high degrees of circuit connectivity. For example, long regulatory regions and high recombination rates promote the evolution of multiple *cis*-regulatory sites by non-adaptive mechanisms (Lynch, 2007a; Ruths and Nakhleh, 2012). Similarly, the greater the permissible degeneracy of *cis*-regulatory sequences, the greater is the probability of multiple connections (Paixão and Azevedo, 2010).

Support for these ideas also comes from direct observation of transcription circuits in various species. First, as previously pointed out, independently evolved circuits show similar, high degrees of connectivity. Recent studies of transcription networks by the Encyclopedia of DNA Elements (ENCODE) project have greatly increased the number of examples where network structure is observed to be highly similar across organisms—in this case, humans, mice, *Caenorhabditis elegans*, and *Arabidopsis thaliana* (Boyle et al., 2014; Stergachis et al., 2014; Sullivan et al., 2014). Although some similarities (e.g., the same master regulator controlling the same biological process in two different species) are clearly conserved from a common ancestor, we argue that similarities in overall network structure are largely due to the pathways we have outlined, in which non-adaptive evolution is a major force.

Second, there are several documented examples in which evolutionary rewiring of an entire network has occurred without apparent changes in the output (Baker et al., 2012; Lavoie et al., 2010; Ludwig et al., 2000; Moses et al., 2006; Schmidt et al., 2010; Tanay et al., 2005; Tsong et al., 2003; 2006). These

studies indicate that, even as the output of a circuit is maintained by stabilizing selection, the individual connections may be free to drift to new configurations.

Third, many connections in networks appear unimportant as assessed by conventional experiments (Fisher et al., 2012; Whitfield et al., 2012). Although it is virtually impossible to prove that a connection is non-functional under all conceivable conditions, several types of experiments suggest that parts of circuits may be functionally unimportant. For example, many direct target genes show no change in transcript levels when a regulator that binds to the gene is deleted or reduced in expression. This behavior describes the majority of the *C. albicans* biofilm network: 60% of binding events do not elicit expression changes when the regulator is deleted, with the provision that biofilms were monitored under a narrow range of conditions. Moreover, many target genes, when deleted, do not appear to compromise the output of the circuit. Although these results can be explained away by circuit compensation, redundancy, inability to monitor a wide variety of conditions, and the like, we suggest it is highly plausible, based on the arguments made above, that many circuit connections simply do not contribute to the output. In any case, many observations made on modern circuits are consistent with a model whereby much of the interconnectedness of transcription circuits has arisen non-adaptively, simply as a consequence of the ease of forming new connections.

Cooperative Binding Produces Connectivity

Cooperative binding is a near-universal feature of eukaryotic transcription regulators, and next we discuss how this property increases the ease of forming new circuit connections and thereby shapes circuit structures. We use the term cooperative binding to mean that the binding of one transcription regulator to a *cis*-regulatory sequence increases the probability that another will occupy a nearby sequence. Mechanistically, this can occur through three distinct means: (1) competitive displacement of nucleosomes, through which binding of one transcription regulator to DNA can increase the accessibility of DNA to a second regulator, thereby increasing its occupancy (Polach and Widom, 1996); (2) a direct, weak, favorable, physical interaction between the two regulators (Johnson et al., 1979); and (3) physical interactions with additional non-DNA-binding proteins that stabilize binding of both of the transcription regulators on DNA (Ptashne and Gann, 2002).

All three forms of cooperative binding would favor the drift of circuits into states of high connectivity by relaxing the *cis*-regulatory sequence requirements needed for a second transcription regulator to be added to a target gene. This idea has an additional implication: cooperativity means that a single change in a *cis*-regulatory sequence or a regulatory protein can establish or eliminate numerous connections. For example, gain of a *cis*-regulatory site for one regulator may allow other regulators to occupy nearby, previously existing weak sites (Figure 4A). Acquisition of a new, favorable protein-protein interaction between transcriptional regulators can have an even more profound effect. Here, cooperative binding can, at least in principle, catalyze the rewiring of an entire set of genes rather than a single gene (Tuch et al., 2008a). In this scenario, the gain of a protein-protein

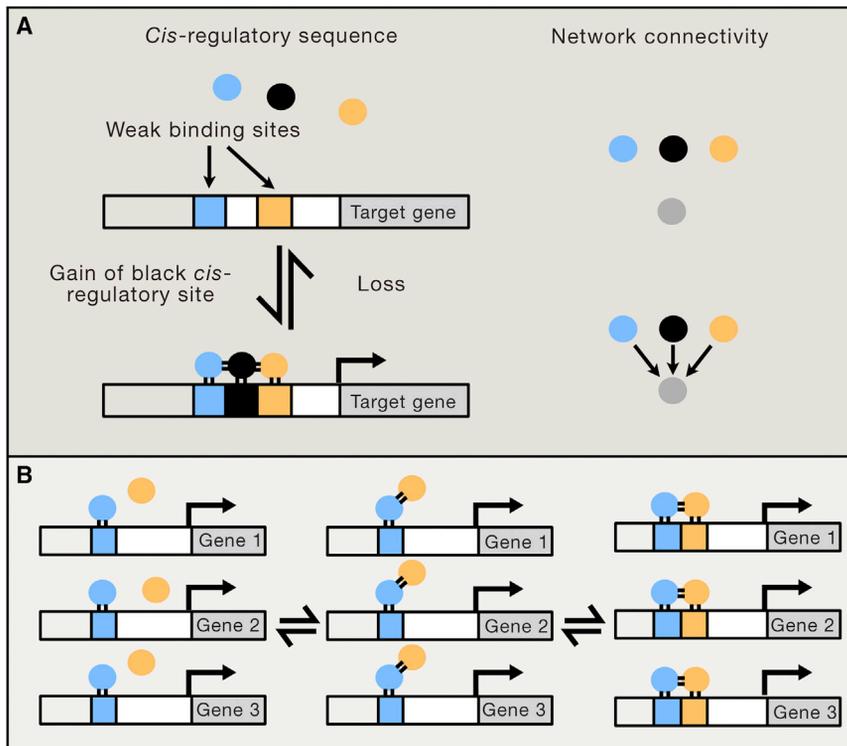


Figure 4. Gain of Multiple Regulatory Connections through Cooperative Binding

(A) Cooperativity between regulators allows binding energy to be shared between protein-DNA and protein-protein interactions. When a strong binding site is gained for one regulator, this may increase the occupancy of regulators on nearby weak binding sites that would otherwise be unoccupied. The effect is a concerted increase in connectivity of that target gene.

(B) The gain of a protein-protein interaction between the blue and orange regulators results in a concerted rewiring of the entire set of genes. As shown in the third panel, direct binding sites for the orange regulator can be gained stepwise at each gene individually without disrupting the circuit. Finally (not shown), the circuit can diversify by moving between equivalent configurations. For example, the ancestral (blue) or derived (orange) connections could be lost without destroying the regulation.

interaction leads to cooperative binding of two regulators when a binding site for only one of the regulators is present (Figure 4B). Following this gain, there can be gene-by-gene gains of *cis*-regulatory sequences for the second regulator. The dually regulated set of genes can then diversify, loosening their connections with the original regulator and strengthening the new ones. In this way, gene sets can be “handed off” from one regulator to another in the course of evolution, a type of change that seems common (Baker et al., 2012; Lavoie et al., 2010; Martchenko et al., 2007; Tanay et al., 2005; Tsong et al., 2006).

The important point is, to influence transcription, transcription regulators must occupy *cis*-regulatory sequence in the cell, but the energy needed for this occupancy can be shared out between protein-DNA and protein-protein interactions. As individual interactions are strengthened and weakened over evolutionary time, the circuit configuration can drift between different “energy-sharing” solutions. Cooperative binding, combined with the ease of strengthening and weakening *cis*-regulatory sequences by random mutations, predicts that networks will drift to high degrees of connectivity—a prediction that is supported experimentally (Baker et al., 2012; Lynch and Wagner, 2008; Stefflova et al., 2013; Tsong et al., 2006). Thus, any two regulators that overlap in their expression would be predicted to share a fraction of their targets under the conditions they are both expressed (Lynch, 2007a), unless these additional connections are specifically selected against.

Formation of Common Network Motifs

One strategy to simplify and understand the function of complex transcription networks has been to search for network motifs—

biology, and it seems likely that, in its most general form (but not necessarily in its detailed mechanism), it is often under purifying selection.

But what about motifs other than positive feedback loops? A more complex network motif known as a feed-forward loop (in which one regulator controls a second regulator and both control the same target gene) is overrepresented in biological networks (Milo et al., 2002). Depending on the parameters of binding, a given feed-forward loop can, in principle, perform logic operations such as pulse detection or expression delay (Alon, 2007; Davidson, 2010). However, it is currently unclear whether the majority of naturally occurring feed-forward loops meet the types of parameter requirements needed for these behaviors.

There are thousands of feed-forward loops embedded in the two networks of Figure 1. We note that feed-forward loops are common byproducts of the evolutionary paths diagrammed in Figures 2 and 3, and thus, the preponderance of feed-forward loops in biological networks may be a result of the same non-adaptive processes that result in large network size and interconnectedness. Indeed, it has been explicitly proposed that many network motifs have arisen as a result of neutral evolutionary processes (Cordero and Hogeweg, 2006; Ingram et al., 2006; Ruths and Nakhleh, 2013; Ward and Thornton, 2007). These ideas contrast with models in which each feed-forward loop in the network possesses optimized parameters that specify a particular transcriptional input-output relationship.

We note that feed-forward loops may also represent non-adaptive intermediates between alternative forms of transcriptional regulation (Figure 5). Rewiring of transcription networks,

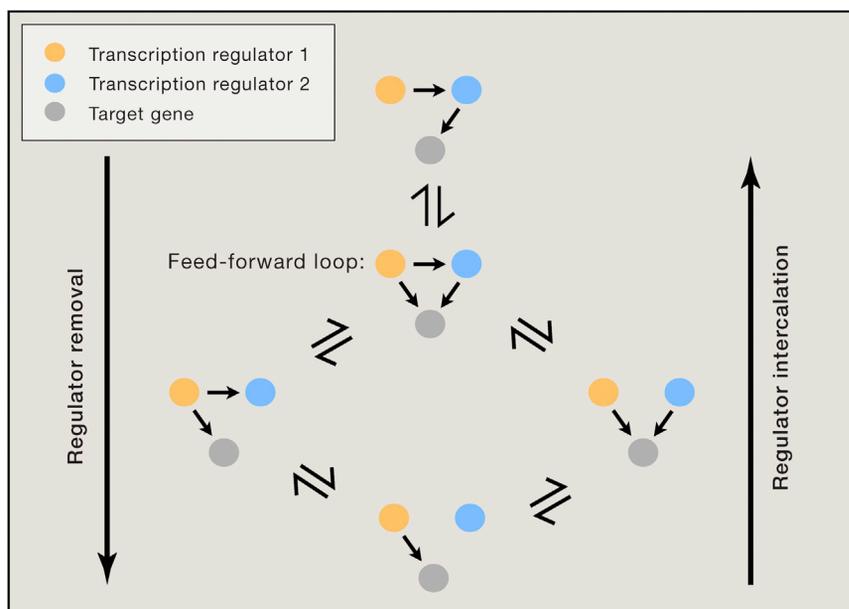


Figure 5. Pathways for Incorporation or Removal of a Transcription Regulator without Breaking the Network

Removal of the blue regulator from the linear regulatory pathway shown in the top network diagram can proceed by first forming a feed-forward loop. Subsequent loss of the connections between the orange and blue regulator and between the blue regulator and the target gene will completely remove the blue regulator from the network as shown in the bottom diagram. The opposite process starting from the bottom diagram and proceeding to the top results in intercalation of the blue regulator into the pathway. If the functions of the blue and orange regulators are redundant in the context of the network, the network can drift between these configurations over evolutionary time without compromising the output of the circuit.

at least in some cases, proceeds through intermediates that are regulated by both the ancestral and derived mechanisms (Li and Johnson, 2010), allowing the regulatory output to be preserved during the rewiring. Although they might arise non-adaptively, feed-forward loops can serve as redundant intermediates between the ancestral and derived states, and thus many observed examples of transcription network rewiring may be a simple consequence of the high frequency with which feed-forward loops are formed by neutral evolution (Lynch, 2007a).

Conclusions

Genomes evolve under selective pressure, but we no longer expect their structures to be orderly, logical affairs dictated by underlying design principles. Here, we have argued that there is no reason to expect transcription circuit networks to be any different. We argue that the drift of transcription networks to steady-state levels of high complexity and interconnection is consistent with the biochemical and biophysical properties of regulatory proteins and *cis*-regulatory sequences, particularly the cooperative binding of transcription regulators to DNA. Combined with universal processes of evolution such as mutation, genetic drift, and selection, network complexity is predicted, from first principles, to be a natural consequence. Complex structures, even if they arose in populations non-adaptively, can nonetheless serve as substrates for future evolutionary innovations or be locked in place by secondary changes; thus, they can be retained by purifying selection even though they arose non-adaptively and are not optimized solutions. If transcription circuits are considered as relatively crude products whose structures are dominated by high-probability evolutionary pathways, many of their more baffling features—size, similarity across diverse species, complexity, redundancy, interconnectedness, for example—begin to make conceptual sense.

ACKNOWLEDGMENTS

We thank C. Nobile, E. Mancera, S. Coyle, and V. Hanson-Smith for helpful comments on the manuscript. We also thank K. Carniol and two anonymous reviewers for providing critical feedback on the content and language. Work from the authors' laboratory was supported by grant R01 GM037049 from the NIH. T.R.S. was supported by a Graduate Research Fellowship from the National Science Foundation.

REFERENCES

- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.
- Baker, C.R., Tuch, B.B., and Johnson, A.D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. USA* 108, 7493–7498.
- Baker, C.R., Booth, L.N., Sorrells, T.R., and Johnson, A.D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* 151, 80–95.
- Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. *Prog. Nucleic Acid Res. Mol. Biol.* 60, 133–168.
- Borneman, A.R., Leigh-Bell, J.A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. (2006). Target hub proteins serve as master regulators of development in yeast. *Genes Dev.* 20, 435–448.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.
- Boyle, A.P., Araya, C.L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L.W., Janette, J., Jiang, L., et al. (2014). Comparative analysis of regulatory information and circuits across distant species. *Nature* 512, 453–456.
- Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36.
- Chan, S.S.-K., and Kyba, M. (2013). What is a master regulator? *J. Stem Cell Res. Ther.* Published online May 4, 2013. <http://dx.doi.org/10.4172/2157-7633.1000e114>.
- Cordero, O.X., and Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. *Mol. Biol. Evol.* 23, 1931–1936.

- Covello, P.S., and Gray, M.W. (1993). On the evolution of RNA editing. *Trends Genet.* 9, 265–268.
- Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920.
- Doolittle, W.F. (2013). Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. USA* 110, 5294–5300.
- Featherstone, D.E., and Broadie, K. (2002). Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *BioEssays* 24, 267–274.
- Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in *Drosophila*. *Nature* 332, 853–856.
- Fisher, W.W., Li, J.J., Hammonds, A.S., Brown, J.B., Pfeiffer, B.D., Weiszmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J.A., Eisen, M.B., et al. (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 109, 21330–21335.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Frankel, N., Wang, S., and Stern, D.L. (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc. Natl. Acad. Sci. USA* 109, 20975–20979.
- Galgoczy, D.J., Cassidy-Stone, A., Llinás, M., O'Rourke, S.M., Herskowitz, I., DeRisi, J.L., and Johnson, A.D. (2004). Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 101, 18069–18074.
- Gerhart, J., and Kirschner, M. (1997). *Cells, Embryos and Evolution* (Wiley-Blackwell).
- Gray, M.W., Lukes, J., Archibald, J.M., Keeling, P.J., and Doolittle, W.F. (2010). Cell biology. Irremediable complexity? *Science* 330, 920–921.
- Guelzim, N., Bottani, S., Bourgine, P., and Képès, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31, 60–63.
- Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. *Science* 267, 1788–1792.
- Hayes, T.E., Sengupta, P., and Cochran, B.H. (1988). The human *c-fos* serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes Dev.* 2 (12B), 1713–1722.
- Hernday, A.D., Lohse, M.B., Fordyce, P.M., Nobile, C.J., DeRisi, J.L., and Johnson, A.D. (2013). Structure of the transcriptional network controlling white-opaque switching in *Candida albicans*. *Mol. Microbiol.* 90, 22–35.
- Ingram, P.J., Stumpf, M.P.H., and Stark, J. (2006). Network motifs: structure does not determine function. *BMC Genomics* 7, 108.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409, 533–538.
- Johnson, A.D., Meyer, B.J., and Ptashne, M. (1979). Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc. Natl. Acad. Sci. USA* 76, 5061–5065.
- Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E.M. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148, 473–486.
- Kakidani, H., and Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. *Cell* 52, 161–167.
- Kielbasa, S.M., and Vingron, M. (2008). Transcriptional autoregulatory loops are highly conserved in vertebrate evolution. *PLoS ONE* 3, e3210.
- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049–1061.
- Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* 8, e1000329.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.
- Li, H., and Johnson, A.D. (2010). Evolution of transcription networks—lessons from yeasts. *Curr. Biol.* 20, R746–R753.
- Liang, Z., and Biggin, M.D. (1998). *Eve* and *ftz* regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 125, 4471–4482.
- Lin, S., and Riggs, A.D. (1975). The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in prokaryotes and eukaryotes. *Cell* 4, 107–111.
- Little, J.W. (2010). Evolution of complex gene regulatory circuits by addition of refinements. *Curr. Biol.* 20, R724–R734.
- Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564–567.
- Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F., and Gray, M.W. (2011). How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* 63, 528–537.
- Lynch, M. (2007a). The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* 8, 803–813.
- Lynch, M. (2007b). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA* 104 (1), 8597–8604.
- Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62, 2131–2154.
- Lynch, M., Field, M.C., Goodson, H.V., Malik, H.S., Pereira-Leal, J.B., Roos, D.S., Turkewitz, A.P., and Sazer, S. (2014). Evolutionary cell biology: two origins, one objective. *Proc. Natl. Acad. Sci. USA* 111, 16990–16994.
- MacArthur, S., Li, X.-Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keränen, S.V.E., et al. (2009). Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 10, R80.
- Martchenko, M., Levitin, A., Hogues, H., Nantel, A., and Whiteway, M. (2007). Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr. Biol.* 17, 1007–1013.
- Mastick, G.S., McKay, R., Oligino, T., Donovan, K., and López, A.J. (1995). Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics* 139, 349–363.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827.
- Monteiro, A. (2012). Gene regulatory networks reused to build novel traits: co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by *optix* expression. *BioEssays* 34, 181–186.
- Mösch, H.U., and Fink, G.R. (1997). Dissection of filamentous growth by transposon mutagenesis in *Saccharomyces cerevisiae*. *Genetics* 145, 671–684.
- Moses, A.M., Pollard, D.A., Nix, D.A., Iyer, V.N., Li, X.-Y., Biggin, M.D., and Eisen, M.B. (2006). Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput. Biol.* 2, e130.
- Nobile, C.J., Fox, E.P., Nett, J.E., Sorrells, T.R., Mitrovich, Q.M., Hernday, A.D., Tuch, B.B., Andes, D.R., and Johnson, A.D. (2012). A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell* 148, 126–138.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.-Y., Shay, T., et al. (2011).

- Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.
- Paixão, T., and Azevedo, R.B.R. (2010). Redundancy and the evolution of cis-regulatory element multiplicity. *PLoS Comput. Biol.* 6, e1000848.
- Pavlicev, M., and Wagner, G.P. (2012). A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol. Evol.* 27, 316–322.
- Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B., and Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proc. Natl. Acad. Sci. USA* 110, 9571–9576.
- Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* 258, 800–812.
- Ptashne, M., and Gann, A. (2002). *Genes and Signals* (Cold Spring Harbor Laboratory Press).
- Qian, W., Ma, D., Xiao, C., Wang, Z., and Zhang, J. (2012). The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep.* 2, 1399–1410.
- Raff, R.A., and Kaufman, T.C. (1983). *Embryos, Genes, and Evolution* (Macmillan Publishing Company).
- Reece-Hoyes, J.S., Pons, C., Diallo, A., Mori, A., Shrestha, S., Kadreppa, S., Nelson, J., Diprima, S., Dricot, A., Lajoie, B.R., et al. (2013). Extensive rewiring and complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network. *Mol. Cell* 51, 116–127.
- Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333, 1137–1141.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.
- Ruths, T., and Nakhleh, L. (2012). ncDNA and drift drive binding site accumulation. *BMC Evol. Biol.* 12, 159.
- Ruths, T., and Nakhleh, L. (2013). Neutral forces acting on intragenomic variability shape the *Escherichia coli* regulatory network topology. *Proc. Natl. Acad. Sci. USA* 110, 7754–7759.
- Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwaile, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040.
- Schweizer, A., Rupp, S., Taylor, B.N., Rölinghoff, M., and Schröppel, K. (2000). The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in *Candida albicans*. *Mol. Microbiol.* 38, 435–445.
- Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 154, 530–540.
- Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., et al. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515, 365–370.
- Stern, D.L., and Orgogozo, V. (2009). Is genetic evolution predictable? *Science* 323, 746–751.
- Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49, 169–181.
- Stone, J.R., and Wray, G.A. (2001). Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18, 1764–1770.
- Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* 8, 2015–2030.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Tanay, A., Regev, A., and Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl. Acad. Sci. USA* 102, 7203–7208.
- Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* 242, 405–411.
- Tsong, A.E., Miller, M.G., Raisner, R.M., and Johnson, A.D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115, 389–399.
- Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature* 443, 415–420.
- Tuch, B.B., Galgoczy, D.J., Hernday, A.D., Li, H., and Johnson, A.D. (2008a). The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* 6, e38.
- Tuch, B.B., Li, H., and Johnson, A.D. (2008b). Evolution of eukaryotic transcription circuits. *Science* 319, 1797–1799.
- Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct conversion of *C. elegans* germ cells into specific neuron types. *Science* 331, 304–308.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041.
- Wagner, A. (2003). Does selection mold molecular networks? *Sci. STKE* 2003, PE41.
- Wang, D.Y.C., Kumar, S., and Hedges, S.B. (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. Biol. Sci.* 266, 163–171.
- Ward, J.J., and Thornton, J.M. (2007). Evolutionary models for formation of network motifs and modularity in the *Saccharomyces* transcription factor network. *PLoS Comput. Biol.* 3, 1993–2002.
- Weirauch, M.T., and Hughes, T.R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* 26, 66–74.
- Weirauch, M.T., and Hughes, T.R. (2011). A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell. Biochem.* 52, 25–73.
- Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M., and Weng, Z. (2012). Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.* 13, R50.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419.
- Zordan, R.E., Miller, M.G., Galgoczy, D.J., Tuch, B.B., and Johnson, A.D. (2007). Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol.* 5, e256.
- Zuckerandl, E. (1997). Neutral and nonneutral mutations: the creative mix—evolution of complexity in gene interaction systems. *J. Mol. Evol.* 44 (1), S2–S8.